DSJ1&2-PR Exh 568

From: Prachi Gururaj
To: Aishwarya Shukla
CC: Tanmay Sengupta; Manjul Kumar; Senthoor Raja; Moutushi Das; Neha Malhotra; Vani Bodapati; Sayid Islam
Sent: 1/24/2014 8:53:57 PM
Subject: RE: The raw data along with questions
Attachments: Data Strategy.docx

Team,
Here are the clarifications:

To begin with, we should be working ONLY with the 'oxycodone' data set that has been provided. The 87 drugs list was just an fyi. Business does not have order size information and other relevant data for these 87 drugs. Hence, we stick to the oxycodone sample dataset for the purpose of our analysis.

1. Order size is measured in 'number of pills'
2. Interpretation of transit time – It is the delivery cycle time of the DCs
3. Retail price – No explanation received, Business not aware if it is price/pill or price/pack.
4. Previous 106 info currently unavailable with Logistics team.
5. Missing number of weeks – It is a result of data pull. Retail link timed out in between during the data pull and hence few weeks are missing
6. Limitation on the number of stores – Simply because it is a sample size considered. Another possible reason being, not all stores carry all the drugs.
7. Prime size – Business unaware
8. Computation logic – Please find attached the methodology documented by Kristy.
   Explanation for "1/100 of 1% of total order size"- In the sample, business intuitively identified the outlier order sizes (90,000, 10,000 etc..) and also decided (from a business sense) that order size 5000 should be a reasonable large cap. All the remaining order sizes (which are above 5000), total upto around 13/14. So, this number '13' just happens to be 1/100 of 1% of the total number of orders placed over the entire time period of 52 weeks. – We should verify this though.

I hope this has answered most of your questions. Do let me know if you need to discuss the computation logic.

Thanks,
Prachi

Prachi Gururaj |+1 (703) 380 5267 | www.mu-sigma.com

From: Aishwarya Shukla
Sent: Friday, January 24, 2014 8:56 AM
To: Prachi Gururaj
Cc: Tanmay Sengupta; Manjul Kumar; Senthoor Raja; Moutushi Das; Neha Malhotra; Vani Bodapati; Sayid Islam
Subject: The raw data along with questions

Hi Prachi,

We have some questions on the data. Please do revert back with clarifications.

\\honts8025\shares\Rx_Compliance\Analytics\Mu_Sigma\Logistics Compliance\Data\Data with questions

Regards,
Aishwarya

Project Goal: To identify and report Suspicious Orders of controlled substances and other frequently abused drugs.

Key Deliverables committed to Focus Area

- Identify "Normal" order levels
- Define "Threshold" and "Hard Limit" levels
- Implement systems enhancements to execute "threshold" and "hard limit" level functions
- Include vendor data in order level limits
- Proactive tracking of order and industry trends
- Plan for continuous update and improvement

Purpose

- Design & operate a system to detect suspicious orders and report to DEA when discovered
- Establish additional maximum order limits of highly abused drugs

Approach
For initial Order Alert upload:
Use 52 week order history to establish "normal" order amounts.
Use Standard Deviation calculations to identify order levels outside the typically expected range.
Provide Reddwerks/KNAPP a data file with the requested Order Alert level for each drug/store combination.
Set Reddwerks/KNAPP Order Alerts that will identify orders outside the typically expected range.
These orders will require investigation and resolution.
For on-going Order Alert evaluation:
Track Order Alerts and investigations.
DC facilities can update individual Order Alert levels as needed.

Data Steps

1. Pulled order shipment information from Retail Link. Information includes order shipment levels for Oxy/APAP 5/325 for the past 52 weeks for every store/club.
2. Calculated the average weekly order volume amount for each store/club
3. Calculated the standard deviation for each order average
4. Added the standard deviation to the average, and identified every order over that number to determine how many Order Alerts would have been generated from this threshold level
    a. 1 Std Dev = +8000 Order Alerts in 52 weeks
    b. 2 Std Dev = +6000 Order Alerts in 52 weeks
    c. 3 Std Dev = +1800 Order Alerts in 52 weeks
5. Stores that don't order often (or new stores) generate a very low threshold, and therefore trigger "false" alerts. To fix this, we put a Minimum threshold on all Order Alerts.
    a. For stores that fall under the minimum threshold, we replace their individual average with the average for all stores (300).

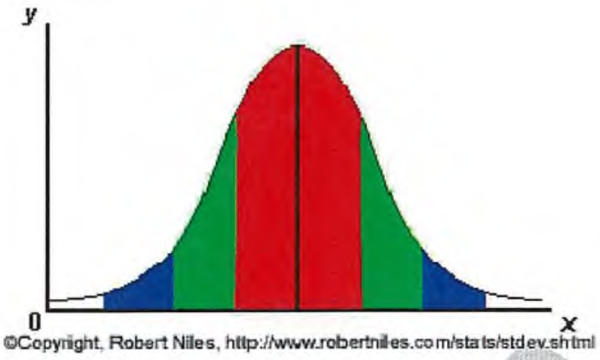
New Order Alerts will be triggered based on the following:

1. Any order greater than 3 Standard Deviations from the average of all store orders will be flagged.
2. If an individual store's volume or volatility require a different alert level, they will alert based on 3 Standard Deviations of their own average.
    a. If that store's average is more than 3 Standard Deviations from the average of all store orders + the average for all stores, the alert will be based on that store's average + 3 Standard Deviations from their average.
3. The data includes some large anomalies, which prevents us from only relying on the Standard Deviation + average calculations.

a. To be sure we are identifying every potentially unusual order, we sorted the averages from largest to smallest. There is an obvious barrier in the data identifying the top 1/100 of 1% (anything over 50 bottles for this particular product).
b. Therefore, any order over 50 bottles for this particular product would also generate an alert.

Number of Order Alerts triggered in 52 week period based on these levels: 158

What is a Standard Deviation?



One standard deviation away from the mean in either direction on the horizontal axis (the two shaded areas closest to the center axis on the above graph) accounts for somewhere around 68 percent of the people in this group. Two standard deviations away from the mean (the four areas closest to the center areas) account for roughly 95 percent of the people. And three standard deviations (all the shaded areas) account for about 99 percent of the people.

If this curve were flatter and more spread out, the standard deviation would have to be larger in order to account for those 68 percent or so of the people. So that's why the standard deviation can tell you how spread out the examples in a set are from the mean.